

LAW OFFICES
WHITHAM, CURTIS & WHITHAM
INTELLECTUAL PROPERTY LAW
11800 SUNRISE VALLEY DRIVE, SUITE 900
RESTON, VIRGINIA 20191

**APPLICATION
FOR
UNITED STATES
LETTERS PATENT**

Applicants: Yoshihiro ONO
For: VOICE TRANSCEIVER
Docket No.: OSP-8180

09315058-012099

VOICE TRANSCEIVER

BACKGROUND OF THE INVENTION

5

Field of the Invention

The present invention relates to a voice transceiver for use in voice transmission by means of digital voice signals which utilize compressed voice coding.

10 This application is based on Patent Application No. Hei 10-144734 filed in Japan, the content of which is incorporated herein by reference.

Relevant Art

Fig. 3 is a block diagram showing the structure of a conventional voice transceiver. In the figure, a decoding code buffer 301 receives compressed voice codes from a circuit (not shown in the figure), and stores these codes into an internal memory. A voice decoder 401 then digitalizes and expands these compressed voice codes stored in the memory of the decoding code buffer 301 into digital voice data.

An SP (speaker) output buffer 501 inputs and stores the voice data expanded by means of the voice decoder 401. D/A converter 601 converts the digital voice data stored in the SP output buffer 501 into an analog voice signal. An amplifier 701 amplifies the analog signal by a predetermined magnitude (amplification), and a speaker 801 then emits the amplified analog voice signal into the air.

In addition, a microphone 802 (hereinafter also referred to as "mic") collects a transmitted voice and converts it into an electronic signal. The microphone 802 then converts the aforementioned conversion result into an analog voice input signal. An amplifier 702 amplifies the analog voice input signal to be inputted by a predetermined magnitude (amplification). A/D converter 602 then converts the analog voice input signal into a digital voice input signal. MIC (microphone) input buffer 502 subsequently stores this digital voice input signal.

6 In 5B' >
ins
10 A voice encoder 401 encodes the digital voice input signal stored in the MIC input buffer 502, and outputs a compressed voice code as a result of the encoding. A compression code buffer 301 then stores the compressed voice code inputted from the voice encoder 401.

In the following, an operation of the voice transceiver according to aforementioned conventional example will be described.

For example, the decoding code buffer 301 temporarily stores the compressed voice code inputted from a communication circuit (not shown in the figures) into an internal memory portion. Subsequently, using the storage of the compressed voice code into decoding code buffer 301 as a trigger, the voice decoder 401 begins processing by expanding the compressed voice code stored in the memory portion of the decoding code buffer 301, and generating digitalized digital voice data.

20 In this manner, the generated voice data is inputted and written into the SP output buffer 501. On the other hand, a voice encoder 402 detects the writing of the digital voice data required for encoding one frame in MIC input buffer 502. The voice encoder 402 then

09315058-052099
660250" 95057E60

commences operation by compressing the digital voice data, and generating a compressed voice code.

Ins B²) Following completion of this operation, the voice encoder 402 outputs the generated compressed voice code to a compression code buffer 602. This compression code buffer
5 602 then stores the inputted compressed voice code. In this manner, the compression code buffer 602 transmits the compressed voice code stored therein to the communication circuit side (not shown in the figures).

Furthermore, the right side of the operation from the D/A converter 601 and A/D converter 602, respectively shown in Fig. 3, is performed during a fixed clock cycle by means of hardware. In other words, the digital voice data of SP output buffer 501 is
10 outputted one sample at a time when necessary, and converted into an analog voice signal by means of D/A converter 601. In addition, at the same time, the analog voice signal inputted from the microphone 802 is sampled when necessary during a fixed cycle, converted into a digital voice signal, and written into MIC input buffer 502 as necessary.

15 However, in the case when the aforementioned voice transceiver is operated in an environment in which the digital voice input signal, which serves as the reception decoding code, is not smoothly and regularly supplied, problems arise such as the generation of interruptions in the output voice from speaker 801, leading to an extreme degradation of the
20 quality of the voice reception (receiving voice quality).

For example, in a personal computer, the supply of a smooth and regular reception signal code cannot be guaranteed by controlling the processing assignment of the processor, by means of processing of the aforementioned voice transceiver (using the same processor)

09315058-052099
660250-85051660

and simultaneously operating an optional user software. As a result, as described above, extreme degradation of the voice reception quality results, when embodying the processing of the aforementioned voice transceiver as software of a personal computer, using a voice transmitter, desktop conference system or the like, which utilizes a personal computer.

5 In addition, in a multi-media transmission terminal, besides voice codes, other data such as images and the like are mixed therein and transmitted. As a result, when the transmission data in a communication circuit is disrupted, it is not possible to specify the disrupted data as voice data or otherwise, and thus a regular and unimpaired supply of reception decoding codes cannot be guaranteed. Consequently, the quality of voice
10 reception is notably degraded even in the voice transmission processing portion of a multi-media transmission terminal.

15 Here, enlargement of the SP output buffer and absorption of the jitter from the output voice may be considered, as an example of a method for avoiding the degradation of the voice reception quality occurring in a voice transmission processing portion of a multi-media transmission terminal. However, an increase in the SP output buffer causes a increase in the shift distance over which the digital voice input signal must pass from the point of input to the point of output. This aspect, in turn, leads to a delay in the voice, and is hence undesirable from a practical standpoint.

In addition, the jitter amount is statistically distributed. As a result, there is a distinct
20 disadvantage in that it is not possible to calculate an absolute value with respect to the optimal amount for enlarging the SP output buffer, as this value changes depending on various conditions.

09315058-052099
660250-8505TE60

Consequently, as a result of the reception voice signal not being supplied after monitoring the remaining data of the SP output buffer, the conventional technology poses problems in that in an environment in which the supply of the reception decoding code is performed in a "burst transmission" manner, when the supply of the receiving decoding code is interrupted, or alternatively when the supply of the receiving decoding code is continued after such an interruption, the SP output voice is similarly interrupted and non-continuous, thereby leading to extreme degradation of the voice reception quality.

SUMMARY OF THE INVENTION

10 In consideration of the aforementioned, it is an object of the present invention to provide a voice transceiver capable of improving the detection performance in the case when the transmission data is disrupted, increasing the reliability of the voice data reception, improving the voice reception quality, and exerting control such that the output voice from the speaker remains smooth and uninterrupted.

15 In order to achieve the aforementioned, the present invention provides, according to a first aspect, a voice transceiver characterized in comprising:

an input means for inputting compressed voice codes of analog data;

an expansion means for digitalizing said compressed voice codes, and expanding and outputting said digital voice data;

20 a buffer means for storing said digital voice data;

a detection means for detecting the quantity of data in said digital voice data stored in said buffer, and outputting a detection signal as a detection result;

a conversion means for converting said digital voice data into analog voice data based on said detection signal; and

a speaker means for emitting said analog voice data into the air.

5 Furthermore, according to a second aspect of the present invention, a voice transceiver is provided which is characterized in further comprising a data control means for controlling the output of said digital voice data to said conversion means, based on said detection signal; wherein, said data control means outputs a dummy code to said expansion means, in the case when said digital voice data stored in said buffer means is less than a
10 required amount for play back; in contrast, in the case when said buffer means approaches an overflow amount, said data control means does not allow the output of said digital voice data to said conversion means.

Moreover, according to a third aspect of the present invention, a voice transceiver is
15 provided, wherein when said dummy code is inputted into said expansion means, said expansion means outputs digital voice data in which the strength of said compressed voice code inputted immediately prior to said dummy signal is reduced.

In addition, according to a fourth aspect of the present invention, a voice transceiver
20 is provided which is characterized in further comprising:

a microphone means for inputting voice data;

a second conversion means for converting said voice data into a digital signal, and outputting this conversion result as other digital voice data; and

an echo component removal means for removing the echo component contained in said other digital voice data.

The voice transceiver according to the present invention possesses a means for
5 inserting a dummy code into the decoding code buffer 301 when the remaining data amount of the digital voice data stored in the SP output buffer 501 becomes small, and a means for discarding the output voice of voice decoder 401 when the remaining data amount of the digital voice data stored in the SP output buffer 501 becomes large (denoted by reference numerals 100 and 200 in Fig. 2).

10 As a result, according to the voice transceiver of the present invention, the voice data is controlled such that underflow of the SP output buffer 501 does not occur, and thus, the speaker output voice remains continuous (i.e., such that a non-continuous speaker output voice does not occur).

In addition, according to the voice transceiver of the present invention, the voice
15 data is also controlled such that overflow of the SP output buffer 501 does not occur, and thus accumulation of delays in the speaker output voice, from the time of input from the transmission source terminal similarly do not occur.

BRIEF EXPLANATION OF THE DRAWINGS

20 Fig. 1 is a block diagram showing the structure of a voice transceiver according to a first embodiment of the present invention.

Fig. 2 is a block diagram showing the structure of a voice transceiver according to a second embodiment of the present invention.

Fig. 3 is a block diagram showing the structure of a conventional example of a voice transceiver.

DESCRIPTION OF THE PREFERRED EMBODIMENTS OF THE INVENTION

5 In the following, the embodiments of the present invention will be described in detail with reference to the figures. Fig. 1 is a block diagram showing the structure of a voice transceiver according to a first embodiment of the present invention. The structure of the transmitting side of the voice transceiver according to this first embodiment, however, is similar to that of the conventional example, and hence its description will be omitted.

10 In the figure, a decoding code buffer 301 is provided which receives and stores a compressed voice code received from the circuit side (not shown in the figures). A voice decoder 401 is also provided which expands the compressed voice code of the aforementioned decoding code buffer 301 into digitalized digital voice data. A selective disposal unit 200 selectively discards digital voice data inputted from this voice decoder 401.

15 SP output buffer 501 stores the digital voice data expanded by means of the voice decoder 401 and discarded by means of the selective disposal unit 200. An insertion/disposal control unit 100 monitors the remaining data amount of the digital voice data stored in SP output buffer 501. In addition, the insertion/disposal control unit 100 outputs both a dummy compression code to the decoding code buffer 301, and a discard request signal to the selective disposal unit 200.

D/A converter 601 converts voice data inputted from SP output buffer 501 into an analog voice signal, and outputs this signal to an amplifier 701. This amplifier 701 amplifies the analog voice signal inputted from D/A converter 601, and outputs the

09315058-052099

amplified signal to a speaker 801. Subsequently, the speaker 801 emits the amplified analog voice signal inputted from the amplifier 701 into the surrounding air.

In the following, an operational example of the voice transceiver according to the aforementioned first embodiment will be described with reference to Fig. 1.

5 For example, when a compressed voice code is received from the communication circuit side (not shown in the figures), the decoding code buffer 301 temporarily stores the inputted compressed voice code into an internal memory portion. Using the writing of the compressed voice code into the aforementioned decoding code buffer 301 as a trigger, the voice decoder 401 then expands the compressed voice code stored in the memory portion of
10 decoding code buffer 301, and generates digitalized digital voice data.

Subsequently, the voice decoder 401 outputs the generated digital voice data to the selective disposal unit 200. If the selective disposal unit 200 has not received a discard request from insertion/disposal control unit 100, then the selective disposal unit 200 writes the supplied digital voice data into SP output buffer 501.

15 On the other hand, if the selective disposal unit 200 has received a discard request from insertion/disposal control unit 100, then the corresponding digital voice data supplied thereto is discarded, and this digital voice data is not written into SP output buffer 501.

Subsequently, the insertion/disposal control unit 100 monitors the remaining data amount of the digital voice data stored in SP output buffer 501, and when this amount falls
20 below a previously set first threshold value, outputs a dummy voice code to the decoding code buffer 301. This first threshold value represents the lower limit (value) of the data amount stored in SP output buffer 501 at which interruption of the output voice does not occur in the speaker 801.

Thereafter, the supply of the dummy voice code serves as a trigger by means of which the operation for generating the aforementioned SP output voice is started. The voice data accumulates in SP output buffer 501 until the data amount stored in SP output buffer 501 exceeds the aforementioned first threshold value.

5 On the other hand, the insertion/disposal control unit 100 monitors the remaining data in SP output buffer 501, and when the remaining data amount exceeds a second threshold value, the insertion/disposal control unit 100 issues a discard request to the selective disposal unit 200. Subsequently, the selective disposal unit 200 conducts the discard (disposal) processing of the digital voice data with regard to which it has received a
10 discard request, wherein the supply of voice data is halted until the data amount returns to below the previously set, second threshold value of SP output buffer 501. This second threshold value represents the upper limit (value) of the data amount which is capable of being stored in SP output buffer 501.

In the following, an applied example based on the first embodiment of the present
15 invention will be described. According to this applied example, a voice transceiver is provided in which the voice codec (voice encoder 402 and decoder 401) conforms to the recommendations of the *ITU-T (International Telecommunication Union) G.723.1*. The structure of the applied example is achieved by means of a voice decoder in which the voice
decoder 401 shown in Fig. 1 of the first embodiment conforms to *ITU-T G.723.1*. Thus,
20 numeral 401 in the same figure may be replaced by a voice decoder 401 which conforms to *ITU-T G.723.1*, which will be described hereafter.

In the following, the operation of the applied example will be described. Initially, upon receiving a voice code conforming to *ITU-T G.723.1* from the communication circuit

09315058-052099

side (not shown in the figure), the decoding code buffer 301 temporarily stores a compressed voice code into an internal memory portion therein. Using the writing of the compressed voice code conforming to *ITU-T G.723.1* into the decoding code buffer 301 as a trigger, the voice decoder 401 conforming to *ITU-T G.723.1* expands the compressed voice code within the decoding code buffer 301, and generates digitalized digital voice data.

The voice codec conforming to *ITU-T G.723.1* is a codec which compresses/decodes 30 msec (240 samples using 16 bit data sampled at 8 kHz) digital voice data as one frame. Consequently, the voice decoder 401 conforming to *ITU-T G.723.1* outputs a voice data frame of 30 msec.

10 The generated voice data frame is then supplied to the selective disposal unit 200. If selective disposal unit 200 has not received a discard request from the insertion/disposal control unit 100, then the supplied voice data frame is written into SP output buffer 501. On the other hand, if selective disposal unit 200 has received a discard request from the insertion/disposal control unit 100, then the supplied voice data frame is discarded without
15 being written into SP output buffer 501.

For example, in the case when the remaining data amount of SP output buffer 501 drops below a first threshold value, the insertion/disposal control unit 100 supplies, at this time, a dummy voice code to the decoding code buffer 301. The dummy voice code supplied in this manner uses a CRC (Cyclic Redundancy Check) error signal supplied at the
20 time when the code is disrupted at the communication circuit.

Subsequently, using the writing of the dummy voice code into the decoding code buffer 301 as a trigger, the voice decoder 401 conforming to *ITU-T G.723.1* performs the decoding process once again. Once the CRC error signal is inputted, the voice decoder 401

09315058-052099
660250" 9505TE60

conforming to *ITU-T G.723.1* generates digital voice data in which the voice of the previous frame has been smoothly reduced.

Consequently, it is possible to smoothly reduce the voice output that is outputted from the speaker 801 before the sound breakup, due to under flow caused by a drop below
5 the first threshold value of SP output buffer 501, occurs.

In the following, a situation in which the remaining data amount of SP output buffer 501 exceeds a second threshold value will be described. At the time when the remaining data amount of SP output buffer 501 exceeds a second threshold value, the
10 insertion/disposal control unit 100 outputs a discard request to the selective disposal unit 200. As a result, the selective disposal unit 200 discards the voice data frame that is outputted from the voice decoder 401 conforming to *ITU-T G.723.1*. In this manner, the insertion/disposal control unit 100 does not allow the supply of digital voice data to occur until the data amount of SP output buffer 501 drops below the second threshold value.

15 In the aforementioned, the first embodiment of the present invention has been described in detail with reference to the figures. However, the concrete structures are not limited to this embodiment, and design modifications are possible within the scope of the present invention, as long as they do not deviate from the essential elements of the present invention.

20 For example, a second embodiment of the present invention will be described with reference to Fig. 2. Fig. 2 is a block diagram showing the structure of a voice transceiver according to a second embodiment.

09315058-052099

In the figure, a decoding code buffer 301 receives a compressed voice code received from the circuit side (not shown in the figure), and stores this code into an internal memory portion therein. The voice decoder 401 expands the compressed voice code inputted from decoding code buffer 301 into digitalized digital voice data. A selective disposal unit 200 is provided for selectively discarding digital voice data inputted from the voice decoder 401.

SP output buffer 501 stores the digital voice data that has been expanded by means of the voice decoder 401, but not been discarded by the selective disposal unit 200. Data which is identical to that supplied to SP output buffer 501 is then supplied and stored in a reference input signal buffer 901. The insertion/disposal control unit 100 monitors the remaining data amount of the digital voice data stored in the reference input signal buffer 901, and respectively supplies a dummy voice code to the decoding code buffer 301, and a discard request signal to the selective disposal unit 200.

D/A converter 601 converts the digital voice data of SP output buffer 501 into an analog voice signal. An amplifier 701 amplifies the analog voice signal inputted from the D/A converter 601. A speaker 801 then emits the amplified analog voice signal into the air.

A microphone 802 is provided for collecting and converting a transmitted voice into an analog voice input signal. An amplifier 702 amplifies the analog voice input signal that is inputted from the microphone 802. A/D converter 602 converts the analog voice input signal that is inputted from amplifier 702 into a digital input signal. MIC input buffer 502 then stores the digitalized digital input signal.

An acoustic echo canceller 902 suppresses the acoustic echo component in the digital input signal. In addition, a voice encoder 402 encodes the digital voice output signal that is outputted from the acoustic echo canceller 902, and outputs the result as a

09315050-062099

EX

compressed voice code. Additionally, a compression code buffer 302 stores the compressed voice code outputted from the voice encoder 402.

In the following, an operational example of the voice transceiver according to the aforementioned second embodiment will be described with reference to Fig. 2.

5 When a compressed voice code is received, for example, from the communication circuit side (not shown in the figure), the decoding code buffer 301 temporarily stores this compressed voice code into an internal memory portion therein. Using the writing of the compressed voice code into the aforementioned decoding code buffer 301 as a trigger, a voice decoder 401 then expands the compressed voice code stored in the memory portion of
10 the decoding code buffer 301, and generates digitalized digital voice data.

Subsequently, the voice decoder 401 outputs the generated digital voice data to a selective disposal unit 200. If the selective disposal unit 200 has not received a discard request signal from the insertion/disposal control unit 100, then the selective disposal unit 200 respectively outputs the supplied digital voice data to SP output buffer 501 and
15 reference input signal buffer 901.

On the other hand, if the selective disposal unit 200 has received a discard request from the insertion/disposal control unit 100, then the corresponding digital voice data supplied thereto is discarded, and this digital voice data is not outputted to SP output buffer 501 or reference input signal buffer 901. In this manner, SP output buffer 501 and reference
20 input signal buffer 901 store the digital voice data supplied from the selective disposal unit 200.

Subsequently, an acoustic echo canceller 902 references the digital voice data stored in the reference input signal buffer 901, and suppresses the echo component in the digital

input signal. The digital voice data of the SP output buffer 501 is then retrieved one sample at a time, as necessary, converted into an analog voice signal in D/A converter 601, and emitted from a speaker 801 after being passing through an amplifier 701.

On the other hand, an analog voice signal inputted from a microphone 802
 5 undergoes sampling, as necessary, by means of A/D converter 602 via amplifier 702. The sampled analog voice signal is then converted into digital input data, and written into MIC input buffer 502. The acoustic echo canceller 902 suppresses the echo component from the digital input data of MIC input buffer 502, and supplies the result to the voice encoder 402.

The voice encoder 402 then encodes the digital input data that is outputted from
 10 acoustic echo canceller 902, and writes the encoded and compressed voice code into a compression code buffer 302. The compressed voice code of the aforementioned compression code buffer 302 is then transferred to the communication circuit side (not shown in the figure).

In addition, the insertion/disposal control unit 100 monitors the remaining data
 15 amount of the digital voice data stored in the reference input signal buffer 901, and when this amount falls below a previously set first threshold value, outputs a dummy voice code to the decoding code buffer 301. Thereafter, the supply of the dummy voice code serves as a trigger by means of which the operation for generating the aforementioned SP output voice is started. The voice data accumulates in the reference input signal buffer 901 until
 20 the data amount stored in this signal buffer 901 exceeds the aforementioned first threshold value.

On the other hand, the insertion/disposal control unit 100 monitors the remaining data in reference input signal buffer 901, and when the remaining data amount exceeds a

09315058-052099

second threshold value, the insertion/disposal control unit 100 issues a discard request to the selective disposal unit 200. As a result, the selective disposal unit 200 conducts the discard (disposal) processing of the digital voice data with regard to which it has received a discard request, wherein the supply of voice data is not conducted until the data amount drops

5 below the second threshold value of the reference input signal buffer 901.

Therefore, as a result of the aforementioned, the voice transceiver according to both the first and second embodiments monitors the data amount of the digital voice data stored in SP output buffer 501, and performs insertion/disposal controls of this digital voice data.

Thus, an effect is obtained that it is possible to smoothly output an output voice from a
10 speaker unit without the occurrence of breakup in the output voice.

Furthermore, according to the voice transceiver of the second embodiment, the operation of an acoustic echo canceller 902 provides the additional effects of stability, since the contents of the output voice of the actual speaker 801 and those of the reference input signal buffer 901 are always in agreement, by means of the insertion/disposal control unit
15 100 which monitors the digital voice data stored in reference input signal buffer 901, and performs the insertion/disposal control of this voice data.

According to a first aspect of the present invention, a voice transceiver is provided which is characterized in comprising: an input means for inputting compressed voice codes
20 of analog data; an expansion means for digitalizing said compressed voice codes, and expanding and outputting said digital voice data; a buffer means for storing said digital voice data; a detection means for detecting the quantity of data of said digital voice data stored in said buffer, and outputting a detection signal as a detection result; a conversion

00315058-052099

means for converting said digital voice data into analog voice data based on said detection signal; and a speaker means for emitting said analog voice data into the air. Therefore, the data amount of the digital voice data stored in said buffer means is monitored, and insertion control of the digital voice data is performed, such that it is possible to smoothly output an output voice from a speaker unit without the occurrence of breakup in the output voice.

According to a second aspect of the present invention, a voice transceiver is provided which is characterized in further comprising a data control means for controlling the output of said digital voice data to said conversion means, based on said detection signal; wherein, said data control means outputs a dummy code to said expansion means, in the case when said digital voice data stored in said buffer means is less than a required amount for play back; in contrast, in the case when said buffer means approaches an overflow amount, said data control means does not allow the output of said digital voice data to said conversion means. As a result, the remaining data amount of the digital voice data stored in said buffer means is monitored, and insertion/disposal control of the digital voice data to the conversion means is performed, such that it is possible to smoothly output an output voice from the actual speaker unit without the occurrence of breakup in the output voice. In addition, the delay can be maintained below a fixed level, since the delays (amounts) from the time of voice input from the transmitting terminal to the speaker output at the "self" terminal do not accumulate.

According to a third aspect of the present invention, a voice transceiver is provided, wherein when said dummy code is inputted into said expansion means, said expansion means outputs digital voice data in which the strength of said compressed voice code inputted immediately prior to said dummy signal is reduced. As a result, the remaining data

amount of the digital voice data stored in said buffer means is monitored, and insertion control of the voice data is performed, such that it is possible to smoothly output an output voice from the actual speaker unit without the occurrence of breakup in the output voice.

According to a fourth aspect of the present invention, a voice transceiver is provided
 5 which is characterized in further comprising: a microphone means for inputting voice data; a second conversion means for converting said voice data into a digital signal, and outputting this conversion result as other digital voice data; and an echo component removal means for removing the echo component contained in said other digital voice data. As a result, by means of monitoring the remaining data amount of the digital voice data stored in
 10 said buffer means of said detection means, and performing insertion/disposal control of the voice data, said echo component removal means provides the additional effects of stability since the contents of the output voice of the actual speaker and those of the buffer means are always in agreement.

09315058-052099